Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

# FEE TRANSMITTAL
## For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

**TOTAL AMOUNT OF PAYMENT** ($) **130.00**

| Complete if Known | |
|---|---|
| Application Number | 10/633,019 |
| Filing Date | July 31, 2003 |
| First Named Inventor | Takeda, Takahiko |
| Examiner Name | Unassigned |
| Art Unit | 2186 |
| Attorney Docket No. | 16869P-085400US |

**METHOD OF PAYMENT** (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 20-1430 Deposit Account Name: Townsend and Townsend and Crew LLP

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

☒ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17 ☒ Credit any overpayments

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038**

**FEE CALCULATION**

**1. BASIC FILING, SEARCH, AND EXAMINATION FEES**

| Application Type | FILING FEES Fee ($) | FILING FEES Small Entity Fee ($) | SEARCH FEES Fee ($) | SEARCH FEES Small Entity Fee ($) | EXAMINATION FEES Fee ($) | EXAMINATION FEES Small Entity Fee ($) | Fees Paid ($) |
|---|---|---|---|---|---|---|---|
| Utility | 300 | 150 | 500 | 250 | 200 | 100 | |
| Design | 200 | 100 | 100 | 50 | 130 | 65 | |
| Plant | 200 | 100 | 300 | 150 | 160 | 80 | |
| Reissue | 300 | 150 | 500 | 250 | 600 | 300 | |
| Provisional | 200 | 100 | 0 | 0 | 0 | 0 | |

**2. EXCESS CLAIM FEES**

| Fee Description | Small Entity Fee ($) | Small Entity Fee ($) |
|---|---|---|
| Each claim over 20 or, for Reissues, each claim over 20 and more than in the original patent | 50 | 25 |
| Each independent claim over 3 or, for Reissues, each independent claim more than in the original patent | 200 | 100 |
| Multiple dependent claims | 360 | 180 |

| Total Claims | Extra Claims | Fee ($) | Fee Paid ($) |
|---|---|---|---|
| _____ | -20 or HP = _____ | x _____ | = _____ |

HP = highest number of total claims paid for, if greater than 20

| Indep. Claims | Extra Claims | Fee ($) | Fee Paid ($) |
|---|---|---|---|
| _____ | -3 or HP = _____ | x _____ | = _____ |

HP = highest number of independent claims paid for, if greater than 3

| Multiple Dependent Claims Fee ($) | Multiple Dependent Claims Fee Paid ($) |
|---|---|
| _____ | _____ |

**3. APPLICATION SIZE FEE**

If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

| Total Sheets | Extra Sheets | Number of each additional 50 or fraction thereof | Fee ($) | Fee Paid ($) |
|---|---|---|---|---|
| _____ - 100 = | _____ / 50 = | _____ (round up to a whole number) x | _____ | = _____ |

**4. OTHER FEE(S)**

| | Fees Paid ($) |
|---|---|
| Non-English Specification, $130 fee (no small entity discount) | |
| Other: Petitions to the Commissioner | **130.00** |

**SUBMITTED BY**

| | | | |
|---|---|---|---|
| Signature | [signature] | Registration No. (Attorney/Agent) 41,405 | Telephone 650-326-2400 |
| Name (Print/Type) | Chun-Pok Leung | | Date January 21, 2005 |

PATENT
Attorney Docket No.: 16869P-085400US
Client Ref. No.: 340200948US1

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re application of:

TAKAHIKO TAKEDA *et al.*

Application No.: 10/633,019

Filed: July 31, 2003

For: STORAGE SYSTEM AND METHOD OF CONTROLLING THE SAME

Customer No.: 20350

Examiner: Unassigned

Technology Center/Art Unit: 2186

Confirmation No.: 4725

**PETITION TO MAKE SPECIAL FOR NEW APPLICATION UNDER M.P.E.P. § 708.02, VIII & 37 C.F.R. § 1.102(d)**

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is a petition to make special the above-identified application under MPEP § 708.02, VIII & 37 C.F.R. § 1.102(d). The application has not received any examination by an Examiner.

(a) The Commissioner is authorized to charge the petition fee of $130 under 37 C.F.R. § 1.17(i) and any other fees associated with this paper to Deposit Account 20-1430.

01/28/2005 MGEBREM1 00000037 201430 10633019
01 FC:1464 130.00 DA

(b) All the claims are believed to be directed to a single invention. If the Office determines that all the claims presented are not obviously directed to a single invention, then Applicants will make an election without traverse as a prerequisite to the grant of special status.

(c) Pre-examination searches were made of U.S. issued patents, including a classification search and a computer database search. The searches were performed on or around July 16, 2004, and were conducted by a professional search firm, Kramer & Amado, P.C. The classification search covered Classes 711 (subclasses 112, 113, 117, 119, 162, and 165) and 714 (subclasses 5, 6, 9, and 47). The computer database search was conducted on the USPTO systems EAST and WEST. The inventors further provided a reference considered most closely related to the subject matter of the present application (see reference #6 below), which were cited in the Information Disclosure Statement filed with the application on July 31, 2003.

(d) The following references, copies of which are attached herewith, are deemed most closely related to the subject matter encompassed by the claims:

(1) U.S. Patent No. 5,720,028;

(2) U.S. Patent No. 6,725,331 B1;

(3) U.S. Patent Publication No. 2003/0105931 A1;

(4) U.S. Patent Publication No. 2003/0221077 A1;

(5) U.S. Patent Publication No. 2004/0123026 A1; and

(6) Japanese Patent Publication No. 10-333838.

(e) Set forth below is a detailed discussion of references which points out with particularity how the claimed subject matter is distinguishable over the references.

A. Claimed Embodiments of the Present Invention

The claimed embodiments relate to a method of configuring pairs each consisting of logical devices to be controlled by two or more storage control apparatus.

Independent claim 1 recites a storage system comprising a host apparatus; and a first storage control apparatus configured to control operations to write data into a storage device serving as a target specified by the host apparatus and read out data from the storage

device. The first storage control apparatus comprises a first processing unit connected to the host apparatus and configured to process a command received from the host apparatus; a cache memory configured to temporarily store data received from the host apparatus; a memory configured to store management information of the storage system; and a second processing unit used configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus.

Independent claim 6 recites, in a storage system coupled to a host apparatus, a first storage control apparatus for controlling operations to write data into a storage device serving as a target specified by the host apparatus and read out data from the storage device. The first storage control apparatus comprises a first processing unit connected to the host apparatus and configured to process a command received from the host apparatus; a cache memory configured to temporarily store data received from the host apparatus; a memory configured to store management information of the storage system; and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connecting the storage system to a second storage control apparatus. When receiving the command, the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus.

Independent claim 8 recites a storage system comprising a first storage control apparatus configured to control operations to read out data from a storage device serving as a target specified by a host apparatus. The first storage control apparatus includes a first storage device configured to store data; a first processing unit connected to the host apparatus and configured to process a read command received from the host apparatus; and a second processing unit configured to read out the data from the first storage device and store the data in a cache memory in accordance with a processing result generated by the first processing unit. The storage system further includes a second storage control apparatus connected to the second processing unit. The second storage control apparatus includes a third processing unit configured to process a read command received from the second processing unit; a second storage device controlled by the second storage control apparatus; and a fourth processing

unit configured to read out the data from the second storage device in accordance with a processing result generated by the third processing unit.

Independent claim 14 recites a control method adopted by a storage system having a host apparatus and a first storage control apparatus for controlling operations to write data into a storage device serving as a target specified by the host apparatus and read out data from the storage device. The first storage control apparatus includes a first processing unit connected to the host apparatus and configured to process a command received from the host apparatus; a first storage device configured to store data specified in a write command received from the host apparatus; a cache memory configured to temporarily store data specified in a write command received from the host apparatus or data read out from the first storage device in accordance with a read command received from the host apparatus; a memory configured to store management information of the storage system; and a second processing unit configured to control an operation to transfer data stored in the cache memory to the first storage device and connected to a second storage control apparatus to control a second storage device. The control method comprises allowing the first processing unit to receive a data write or read command from the host apparatus; determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second storage control apparatus if the step of determining indicates that a command received from the host apparatus has been issued to the logical device on the second storage device.

B. Discussion of the References

None of the references disclose a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus. Nor do they teach that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus.

1. U.S. Patent No. 5,720,028

This reference discloses an external storage system having a storage unit for storing data and a plurality of storage controllers for controlling data transfer between an upper level system and the storage unit, with each storage controller having a controller for controlling the operation of the storage controller, and an external storage system having a management memory for storing management information of a plurality of storage controllers. The external storage system has a first storage controller for processing an input-output request and a second storage controller for standing by and having the process to be executed by the first storage controller and partially executed by the second storage controller.

The reference merely discloses an external storage system having a first storage controller for processing an input-output request and a second storage controller to standing by. It does not teach a storage system having a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus, as recited in claim 1. Nor does it disclose that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus, as recited in claim 6. It also fails to teach a processing unit in a first storage device configured to read out data from the first storage device and store the data in a cache memory in accordance with a processing result generated by another processing unit in the first storage device, and a processing unit in a second storage device configured to read out the data from the second storage device in accordance with a process result generated by another processing unit in the second storage device based on a read command from the processing unit from the first storage device, as recited in claim 8. It further fails to disclose determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second storage control apparatus if the step of determining indicates

that a command received from the host apparatus has been issued to the logical device on the second storage device, as recited in claim 14.

2. U.S. Patent No. 6,725,331 B1

This reference discloses a method and an apparatus for managing the dynamic assignment resources in a data storage system, having a plurality of storage devices, a plurality of controllers that are each coupled to at least one of the plurality of storage devices and controls access to the one of the plurality of storage devices, a memory that is globally accessible to each of the plurality of controllers; first means for creating in the memory a global table that stores information that specifies dynamic assignments of resources in the storage system, and second means for creating a local table in at least one of the plurality of controllers that includes all of the information stored in the global table.

The reference merely discloses a technique for managing the dynamic assignment resources in a data storage system. It does not teach a storage system having a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus, as recited in claim 1. Nor does it disclose that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus, as recited in claim 6. It also fails to teach a processing unit in a first storage device configured to read out data from the first storage device and store the data in a cache memory in accordance with a processing result generated by another processing unit in the first storage device, and a processing unit in a second storage device configured to read out the data from the second storage device in accordance with a process result generated by another processing unit in the second storage device based on a read command from the processing unit from the first storage device, as recited in claim 8. It further fails to disclose determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second

storage control apparatus if the step of determining indicates that a command received from the host apparatus has been issued to the logical device on the second storage device, as recited in claim 14.

3. U.S. Patent Publication No. 2003/0105931 A1

This reference discloses an architecture for transparent mirroring, having a data storage system including receiving a request by a first data storage device controller for data access operation, with data written to a local storage device and a data access operation performed by a second data storage device controller communicatively coupled to the first data storage device controller. The second data storage controller is communicatively coupled to a second data storage device.

The reference merely discloses an architecture for transparent mirroring. It does not teach a storage system having a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus, as recited in claim 1. Nor does it disclose that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus, as recited in claim 6. It also fails to teach a processing unit in a first storage device configured to read out data from the first storage device and store the data in a cache memory in accordance with a processing result generated by another processing unit in the first storage device, and a processing unit in a second storage device configured to read out the data from the second storage device in accordance with a process result generated by another processing unit in the second storage device based on a read command from the processing unit from the first storage device, as recited in claim 8. It further fails to disclose determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second storage control apparatus if the step of determining indicates

that a command received from the host apparatus has been issued to the logical device on the second storage device, as recited in claim 14.

4. U.S. Patent Publication No. 2003/0221077 A1

This reference discloses a method for controlling storage system, and storage control apparatus, including a method for controlling a storage system having a host computer, and a first storage control apparatus and a second storage control apparatus each receiving a data input/output request from the host computer and executing a data input/output process for a storage device in response to the request; connecting a first communication path between the host computer and the first apparatus; connecting a second communication path between the first apparatus and the second apparatus; and receiving a first data input/output request from the host computer through the first path by the first apparatus; when the first apparatus has judged that the first request is not for the first apparatus, transmitting by the first apparatus a second data input/output request corresponding to the first request, to the second apparatus through the second path; and by the second apparatus, receiving the second request and executing a data input/output process corresponding to the second request received.

The reference merely discloses connecting two storage control apparatuses by a second communication path to transfer data input/output request. It does not teach a storage system having a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus, as recited in claim 1. Nor does it disclose that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus, as recited in claim 6. It also fails to teach a processing unit in a first storage device configured to read out data from the first storage device and store the data in a cache memory in accordance with a processing result generated by another processing unit in the first storage device, and a processing unit in a second storage device configured to read out the data from the second storage device in accordance with a process

result generated by another processing unit in the second storage device based on a read command from the processing unit from the first storage device, as recited in claim 8. It further fails to disclose determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second storage control apparatus if the step of determining indicates that a command received from the host apparatus has been issued to the logical device on the second storage device, as recited in claim 14.

5. U.S. Patent Publication No. 2004/0123026 A1

This reference discloses a control method for storage device controller system, and storage device controller system provided with a first storage device controller that is connected to first and second storage devices and that has first and second communications control means that receive data input/output requests from a mainframe computer and an open system computer, respectively, and a second storage device.

The reference merely discloses a control method for an arrangement including a first storage device controller and first and second storage devices. It does not teach a storage system having a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system to a second storage control apparatus, as recited in claim 1. Nor does it disclose that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus, as recited in claim 6. It also fails to teach a processing unit in a first storage device configured to read out data from the first storage device and store the data in a cache memory in accordance with a processing result generated by another processing unit in the first storage device, and a processing unit in a second storage device configured to read out the data from the second storage device in accordance with a process result generated by another processing unit in the second storage device based on a read command from the processing unit from the first storage device, as recited in claim 8. It

further fails to disclose determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second storage control apparatus if the step of determining indicates that a command received from the host apparatus has been issued to the logical device on the second storage device, as recited in claim 14.

6. Japanese Patent Publication No. 10-333838

This reference relates to a data multiplexing storage subsystem to enhance the processing performance of each storage subsystem by suppressing the increase of a load accompanying data copy for data multiplexing between plural storage subsystems. In the storage subsystem, a master disk subsystem 104, which has a constitution connecting disk storage device 116 under the command of a disk controller 103, is connected through an inter-master and remote data transfer path 117 with a remote disk subsystem 105 in the same constitution, and write data outputted from a central processing unit 101 through a data transfer path 102 to the master disk subsystem 104 are copies to the remote disk subsystem 105. In this case, the storage subsystem is provided with a data update under of time storage table 109 which manages a data update history for each data management unit such as a track, cylinder, volume, and file, and the continuity of the data update is judged for each data management unit, and an executing equipment is controlled for summarizing the data copying operation so that a load resulted from the execution of the data copy can be reduced.

The reference merely discloses a remote copy in which an attempt may be made to allow accesses to be made to individual logical devices of a logical-device pair. In such a case, a host apparatus needs to issue a command to make a transition to a state known as a split state. This command causes each of the logical devices composing the pair to transit to a state in which the access history of one of the logical devices is managed separately from the management of the access history of the other logical device. See present application, at page 1, line 18 to page 2, line 8.

However, the reference does not teach a storage system having a first storage control apparatus which includes a first processing unit configured to process a command from a host apparatus and a second processing unit configured to control an operation to transfer data stored in the cache memory to the storage device and connect the storage system

to a second storage control apparatus, as recited in claim 1. Nor does it disclose that the first processing unit references the management information held in the first storage control apparatus to determine whether the command is a command issued to a logical device on a storage device controlled by the first storage control apparatus or a command issued to a logical device on a storage device controlled by the second storage control apparatus, as recited in claim 6. It also fails to teach a processing unit in a first storage device configured to read out data from the first storage device and store the data in a cache memory in accordance with a processing result generated by another processing unit in the first storage device, and a processing unit in a second storage device configured to read out the data from the second storage device in accordance with a process result generated by another processing unit in the second storage device based on a read command from the processing unit from the first storage device, as recited in claim 8. It further fails to disclose determining whether a command received from the host apparatus has been issued to a logical device on the first storage device or a logical device on the second storage device on the basis of the management information; and providing a control command to the second storage control apparatus if the step of determining indicates that a command received from the host apparatus has been issued to the logical device on the second storage device, as recited in claim 14.

(f) In view of this petition, the Examiner is respectfully requested to issue a first Office Action at an early date.

Respectfully submitted,

Chun-Pok Leung
Reg. No. 41,405

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400
Fax: 415-576-0300
Attachments
RL:rl
60297662 v1

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-333838
(43)Date of publication of application : 18.12.1998

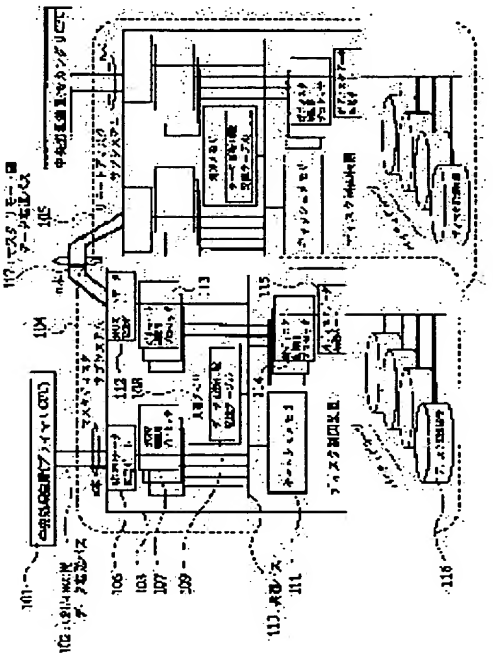
(51)Int.Cl. G06F 3/06

(21)Application number : 09-146652
(22)Date of filing : 04.06.1997

(71)Applicant : HITACHI LTD
(72)Inventor : AZUMI YOSHIHIRO
IZUMI HIROYUKI
NAKANISHI HIROAKI

## (54) DATA MULTIPLEXING STORAGE SUB-SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To enhance the processing performance of each storage sub-system, by suppressing the increase of a load accompanying data copy for data multiplexing between plural storage sub-systems.
SOLUTION: In this storage sub-system, a master disk sub-system 104, which has a constitution connecting disk storage devices 116 under the command of a disk controller 103, is connected through an inter-master and remote data transfer path 117 with a remote disk sub-system 105 in the same constitution, and write data outputted from a central processing unit 101 through a data transfer path 102 to the master disk sub-system 104 are copied to the remote disk sub-system 105. In this case, this storage sub-system is provided with a data update under of time storage table 109 which manages a data update history for each data management unit such as a track, cylinder, volume, and file, and the continuity of the data update is judged for each data management unit, and an executing equipment is controlled for summarizing the data copying operation so that a load resulted from the execution of the data copy can be reduced.



LEGAL STATUS

| | |
|---|---|
| [Date of request for examination] | 13.09.2000 |
| [Date of sending the examiner's decision of rejection] | |
| [Kind of final disposal of application other than the examiner's decision of rejection or application converted registration] | |
| [Date of final disposal for application] | |
| [Patent number] | 3400297 |
| [Date of registration] | 21.02.2003 |
| [Number of appeal against examiner's decision | |

of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開平10－333838
(43)公開日　平成10年(1998)12月18日

| (51)Int.Cl.[6] | 識別記号 | ＦＩ | |
|---|---|---|---|
| Ｇ０６Ｆ　3/06 | ３０４ | Ｇ０６Ｆ　3/06 | ３０４Ｅ |

審査請求　未請求　請求項の数３　ＯＬ　（全 15 頁）

(21)出願番号　特願平9－146652

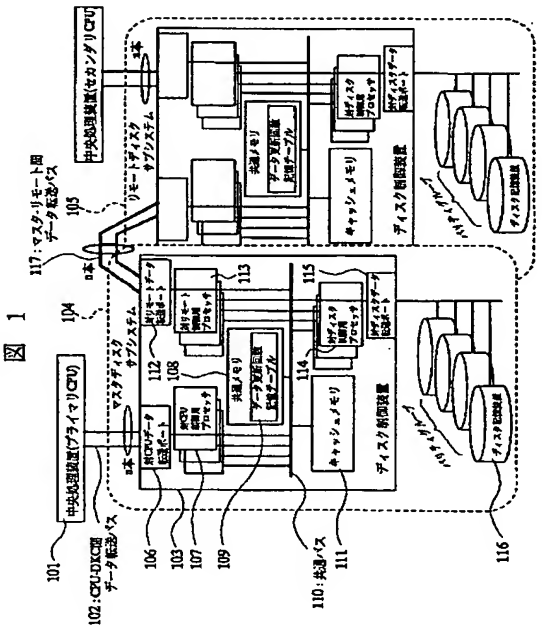
(22)出願日　平成９年(1997)６月４日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　安積　義弘
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(72)発明者　泉　洋行
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(72)発明者　中西　弘晃
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(74)代理人　弁理士　筒井　大和

(54)【発明の名称】　データ多重化記憶サブシステム

(57)【要約】
【課題】　複数の記憶サブシステム間でのデータ多重化のためのデータ複写に伴う負荷の増大を抑制して各記憶サブシステムの処理性能を向上させる。
【解決手段】　ディスク制御装置１０３の配下にディスク記憶装置１１６を接続した構成のマスタディスクサブシステム１０４と、これと同一構成のリモートディスクサブシステム１０５とをマスターリモート間データ転送パス１１７にて接続し、データ転送パス１０２を介して中央処理装置１０１からマスタディスクサブシステム１０４に出力されるライトデータをリモートディスクサブシステム１０５に複写する構成において、トラック、シリンダ、ボリューム、ファイル等のデータ管理単位毎のデータ更新履歴を管理するデータ更新回数記憶テーブル１０９を設け、データ管理単位ごとにデータ更新の継続性を判断してデータ複写操作が集約されるように実行契機を制御して、データ複写の実行に起因する負荷を軽減する。

図１



【特許請求の範囲】

【請求項１】　上位装置と第１のデータ転送経路を介して接続された記憶制御装置および当該記憶制御装置の配下の記憶装置を含む第１の記憶サブシステムと、各々が、記憶制御装置および当該記憶制御装置の配下の記憶装置を含み、第２のデータ転送経路を介して前記第１の記憶サブシステムに接続された少なくとも一つの第２の記憶サブシステムと、からなり、前記第１のデータ転送経路を介して前記上位装置から受領したライトデータを前記第１の記憶サブシステム内の前記記憶装置に格納するとともに、任意の契機にて前記第２のデータ転送経路を介して前記第２の記憶サブシステム内の前記記憶装置に複写することによって、前記第１および第２の記憶サブシステムにて前記ライトデータを多重に保持するデータ多重化記憶サブシステムであって、

前記上位装置から到来する前記ライトデータの前記第１の記憶サブシステムの前記記憶装置における所望のデータ管理単位に対する書き込み履歴を記憶する制御情報記憶手段と、

前記制御情報記憶手段の前記書き込み履歴を参照し、前記上位装置から前記データ管理単位に対する単位時間当たりのデータ書き込み要求の偏りを観測して、前記第２の記憶サブシステムに対する前記ライトデータの前記複写の実行方法および実行契機の少なくとも一方を制御する制御論理と、

を備えたことを特徴とするデータ多重化記憶サブシステム。

【請求項２】　請求項１記載のデータ多重化記憶サブシステムにおいて、

前記書き込み履歴を記憶する前記制御情報記憶手段として複数の記憶テーブルを備え、所望の周期毎に前記記憶テーブルを切替えることで、時間の経過に対する前記上位装置からのデータ書き込み要求の変化を観測し、観測された前記データ書き込み要求の変化に応じて、前記上位装置からの前記データ書き込み要求の継続性を推測し、前記第２の記憶サブシステムに対する前記ライトデータの前記複写の実行契機を制御する第１の制御動作、

前記第１の記憶サブシステムのデータを一括して前記第２の記憶サブシステムに複写する際に、前記制御情報記憶手段に格納されている前記書き込み履歴を参照することによって、複写対象範囲に対する前記上位装置からのデータ書き込み傾向を観測し、前記上位装置からのデータ書き込み要求が継続または集中している未複写の前記データ管理単位に対する前記複写の実行契機を遅らせる第２の制御動作、

の少なくとも一方の制御動作を行うことを特徴とするデータ多重化記憶サブシステム。

【請求項３】　請求項１記載のデータ多重化記憶サブシステムにおいて、前記第１および第２の記憶サブシステムの各々の前記記憶装置は、グループを構成する複数のデータ単位および当該データ単位から生成された冗長データを複数個の記憶媒体に分散して格納する論理的または物理的な冗長記憶構成を備え、

前記書き込み履歴を記憶する前記制御情報記憶手段を参照して個々の前記グループ内の前記データ単位に対するデータ書き込み要求の発生状態を観測し、当該グループ内の全ての前記データ単位を一括して前記第２の記憶サブシステムに転送する第１の複写操作、および前記グループ内の複数の前記データ単位のうち、前記上位装置からのデータ書き込み要求によって更新された前記データ単位のみを選択的に前記第２の記憶サブシステムに転送する第２の複写操作、の選択および実行契機を制御する第３の制御動作、

前記第３の制御動作において、前記第１および第２の前記記憶サブシステム間におけるデータ転送所要時間を含めた前記複写操作の所要時間を観測し、前記第１および第２の複写操作の選択および実行契機の制御を行う第４の制御動作、

の少なくとも一方の制御動作を行うことを特徴とするデータ多重化記憶サブシステム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、データ多重化記憶技術に関し、特に、遠隔地等に独立に分散して配置された複数の記憶サブシステムにて同一データを多重に分散して保持することでデータ保障を実現する情報処理システム等に適用して有効な技術に関する。

【０００２】

【従来の技術】中央処理装置とディスク記憶装置に代表される周辺記憶装置（主にディスク記憶装置とディスク制御装置とからなるディスクサブシステム）とからなる情報処理システムでは、情報量の膨大化とともに取り扱うデータの記憶に対する信頼性への要求が強まる中で、従来よりディスク装置などの記憶媒体や記憶装置の物理的な障害に対する信頼性向上策として、複数個の記憶媒体にデータを二重に保持することによって、障害に伴うデータ消失に対しバックアップデータからの回復を図るデータ二重化記憶サブシステムが実用化されている。また、データを複数個のディスク装置に分割して配置し、更に幾つかのデータを一単位としてパリティーデータに代表される冗長データを作成・記憶することによって、あるデータの媒体障害やディスク装置の障害時に、冗長データと当該一単位内の他データとからデータ回復を行なうＲＡＩＤ記憶装置も実用化されている。

【０００３】ところが銀行等のオンラインシステムに代表されるように、広域に渡って情報処理システムが機能し、多くの情報処理システムが連動しているようなシステムにおいては、これらのデータ信頼性向上技術は、一つの記憶サブシステム内でデータを多重に保持したり冗長化を図るものであり、その記憶サブシステム全体の障

害や、中央処理装置をも含む情報処理システム全体がたとえば建物全体の停電・火災等によって動作しなくなった場合、その被害が広域のシステム全体に影響を及ぼすばかりでなく、データ消失に伴う被害度は甚大なものになってしまう。この様な懸念に対し、遠隔地においてデータを二重に保持するデータ二重化管理システムが実用化されている。しかしながらこの遠隔データ二重化においては、遠隔地に設置された情報処理システム間のデータの通信を中央処理装置間の通信機能によって処理しているため、データ処理や演算等を行なう中央処理装置の負荷が大きく、この中央処理装置の負荷を軽減することが、遠隔データ二重化システムの課題とされている。

【０００４】この様な課題に対し、ディスク制御装置に制御装置間で通信およびデータ転送を行なう機能を設け、遠隔地にあるそれぞれの情報処理システムの制御装置同士を、通信・データ転送パスで接続することにより、データ二重化に掛かる負荷を記憶制御装置に負担させることで中央処理装置の負荷を軽減するシステムも実用化されている。この遠隔データ二重化記憶サブシステムでは、主業務を行なう情報処理システムをプライマリシステムとし、それぞれ第１の中央処理装置、第１のディスクサブシステムおよび第１のディスク制御装置とする。また、バックアップ側の情報処理システムをセカンダリシステムとし、それぞれ第２の中央処理装置、第２のディスクサブシステムおよび第２のディスク制御装置とする。第１、第２のそれぞれのディスク制御装置は不揮発化機構を備えた大容量のキャッシュメモリ（ディスクキャッシュ）を備えている場合が一般的である。第１と第２のディスク制御装置間を１本ないしは複数本のデータ転送パスで接続し、データの一単位毎（たとえばボリューム毎）に正・副のペアボリュームの関係を定義する。正側のデータをマスタデータ（マスタボリューム）呼び、副側データをリモートデータ（リモートボリューム）と呼ぶ。プライマリシステムでのディスクサブシステムへのＷＲＴ＿Ｉ／Ｏにおいては、第１の中央処理装置から第１のディスクサブシステムへの書き込みデータを、自配下のディスク記憶装置に書き込むだけでなく、第２のディスク制御装置のホストとして第２のディスクサブシステムにデータ書き込みＩ／Ｏを発行し、データの二重化を図る。この様にしてデータファイルの二重化の運用を行なっている最中に、プライマリシステム側で障害が発生し、業務の継続が不可能になった場合には、即座にセカンダリシステムに業務を切替え、二重化されている第２のディスクサブシステムのデータを元に業務を継続する。

【０００５】なお、データの二重化技術としては、たとえば、米国特許第５，１５５，８４５号に開示される技術が知られている。この技術では、分散して配置された複数の制御ユニットと、各制御ユニットの配下に等価な構成で接続された記憶手段とを設け、ひとつの制御ユニットがレコードの書き込み要求を受けると他の各制御ユニット配下の対応するすべてのボリュームに当該レコードのコピーが書き込まれるようにしたものである。

【０００６】前述の第１のディスク制御装置からのＷＲＴ＿Ｉ／Ｏによるデータ二重化においては、データ二重化の契機に関し、主に以下の二通りの方式がある。

【０００７】（１）同期方式

プライマリシステムの第１の中央処理装置から第１のディスクサブシステムへのＷＲＴ＿Ｉ／Ｏに同期して、第２のディスクサブシステムに同一データのＷＲＴ＿Ｉ／Ｏを発行することによって、プライマリ側のデータとセカンダリ側のデータが常に同期しているように制御する方式。第１のディスク制御装置は、第１の中央処理装置からのＷＲＴ＿Ｉ／Ｏ時に、自制御装置内のキャッシュメモリにＷＲＴデータを書き込んだ時点で、データ転送の完了報告を行ない、その後第２のディスクサブシステムに同一のＷＲＴ＿Ｉ／Ｏを発行し自キャッシュメモリ上のデータを第２のディスク制御装置に転送することによって、二重化のためのＷＲＴ＿Ｉ／Ｏ処理を行なう。第２のディスクサブシステムへのＷＲＴ＿Ｉ／Ｏが完了した時点で、第１のディスク制御装置は第１の中央処理装置にＩ／Ｏ完了報告を行なう。即ち、第１の中央処理装置がＷＲＴ＿Ｉ／Ｏの完了報告を受領した時点で、第２のディスクサブシステムへのデータ複写は完了しているため、プライマリ側のデータとセカンダリ側のデータの同期性は保たれる。

【０００８】（２）非同期方式

第１のディスクサブシステムのデータ更新に対して、第２のディスクサブシステムへのデータの更新を非同期に行なう方式。第１のディスク制御装置は第１の中央処理装置からのＷＲＴ＿Ｉ／Ｏ時に、第１のディスクサブシステムにデータを書き込んだだけでＩ／Ｏ完了報告を行なう。第１のディスクサブシステムへのデータの書き込み関しては、ディスク記憶装置の記憶媒体へのデータ書き込みが終わってからＩ／Ｏ完了報告としても良いし、ディスク制御装置内のキャッシュメモリにデータを格納しただけでＩ／Ｏ完了報告を行なっても良い。第１のディスクサブシステムに書き込まれたが第２のディスクサブシステムに対してはデータの反映を行なっていないデータは、第１のディスク制御装置にて未反映データとして管理される。第１のディスクディスク制御装置は、一定周期や中央処理装置からのデータ反映要求、もしくは未反映データの残留量に応じて、中央処理装置からのＷＲＴ＿Ｉ／Ｏとは非同期に、第２のディスクサブシステムに対してＷＲＴ＿Ｉ／Ｏを起動し、未反映データの書き込みを行なう。

【０００９】また、既に第１のディスクサブシステム上に存在するデータボリュームを新たに遠隔二重化ボリュームとして定義し二重化ペアを新規に作成する場合（これを初期コピーと呼ぶ）には、第１のディスク制御装置

は、第１のディスクサブシステムの当該ボリュームのデータを順次にディスク記憶装置からキャッシュメモリに読み出し、第１のディスク制御装置から第２のディスクサブシステムに書き込みＩ／Ｏを発行することによって、ボリュームデータの複写を行なう。この時のデータ複写の一単位はデータ格納単位の一単位（トラック）毎であっても良いし、複数個のデータ単位（たとえば、シリンダ）毎であっても構わない。

【００１０】更に、第１のディスクサブシステムは、初期コピー処理のＩ／Ｏを実行しながら、同時に第１の中央処理装置からの更新Ｉ／Ｏを受けることも可能である。初期コピー実行中のボリューム上のデータに対する更新においては、第１のディスク制御装置は、その更新範囲が、初期コピー処理が実施済み（第２のディスクサブシステムへの複写が完了済み）の領域に対する更新の場合には、同期または非同期の方式において第２のディスクサブシステムへの更新データの反映を行なう。また、更新範囲が初期コピー未実施の領域に対する更新の場合には、いずれ初期コピーのための二重化ＷＲＴ＿Ｉ／Ｏによって第２のディスクサブシステムへのデータ複写が行われるので、第１のディスクサブシステムへのデータ更新のみであっても構わない。

【００１１】ところで、第１および第２のディスクサブシステムは、以下に述べるようなＲＡＩＤ－５のデータ格納方式であっても良い。本技術は、Ｄ．Ａ，Ｐａｔｔｅｒｓｏｎ，ｅｔ，ａｌ．“Ｉｎｔｒｏｄｕｃｔｉｏｎ ｔｏ ＲｅｄｕｎｄａｎｔＡｒｒａｙｓ ｏｆ Ｉｎｅｘｐｅｎｓｉｖｅ Ｄｉｓｋｓ（ＲＡＩＤ）”，ｓｐｒｉｎｇ ＣＯＭＰＣＯＮ’８９，ｐｐ．１１２－１１７，Ｆｅｂ．１９８９の論文にて述べられている技術である。ＲＡＩＤ－５とは、ディスクサブシステムをｎ＋ｍ個のディスク記憶装置を一つのデータ格納単位とし、データのある一単位（たとえば、ディスク媒体上の１トラック）毎に、ｎ個のディスク記憶装置に分割して格納する。さらにｎ個のデータ単位を１グループとしてパリティデータと呼ばれる冗長データを作成する。冗長データ数はその冗長度に応じて定まり、冗長度がｍの場合はｍ個の冗長データを作成する。冗長データそのものも当該冗長データを構成するデータグループの格納ディスク装置とはまた異なるディスク装置に格納する。このｎ個のデータ単位とそのｍ個の冗長データから構成されるデータ群を冗長化グループと呼ぶ。このことにより、一つのディスク記憶装置が障害により読み出し不能に陥ったとしても、当該冗長化グループの他のｎ－１個のデータとｍ個の冗長データからデータの再生が可能であり、また同様に障害によって書き込み不良に陥った場合でもｍ個の冗長データを更新しておくことで論理的にデータの格納がなされる。このようにしてディスク装置やディスク媒体の障害に対しデータの信頼性を高めている。さて、ＲＡＩＤ－５のデータ記憶方式においては、データの更新に際し、主に以下の２通りの冗長データ作成方法がある。

【００１２】（１）全ストライプライト方式
冗長化グループを構成するデータ単位グループをストライプ列と定義し、これらの全データ単位から冗長データを新たに作り出す方式。

【００１３】（２）リードモディファイライト方式
冗長化グループを構成するデータ単位のある一単位が更新された場合に、更新データ単位の旧データと更新データと旧冗長データとを演算し、新冗長データを作成する方式。中央処理装置からのある一単位データの更新時に、ディスク制御装置はキャッシュメモリ上に旧データと新データを保持し、また当該冗長化グループの冗長データがキャッシュメモリ上に存在しない場合には、冗長化データをディスク記憶装置からキャッシュメモリ上に読み出し、新冗長データを作成する。この様にデーター単位の更新に対し余分に冗長データのディスク装置からの読み出し・書き込みが発生することをライトペナルティと呼ぶ。

【００１４】
【発明が解決しようとする課題】上述の遠隔データ二重化においては、同期方式の二重化を採用した場合、第１の中央処理装置からのＷＲＴ＿Ｉ／Ｏ時の応答時間は、Ｉ／Ｏ完了報告前に第２のディスクサブシステムへのデータ書き込みＩ／Ｏを行なうために、約二倍の処理時間が必要となる。また、非同期方式の二重化を採用した場合においても、中央処理装置からのＷＲＴ＿Ｉ／Ｏの応答時間そのものは維持されるものの、ディスク制御装置のスループットは二重化のためのＩ／Ｏ処理の負荷により劣化は免れ得ない。このため、遠隔二重化のシステムにおいては、ディスク制御装置のデータ二重化処理の効率向上が性能上の最大の技術的課題となる。

【００１５】中央処理装置からのボリュームデータの更新処理は、その形態によってはある特定の領域にアクセスが集中し、たとえばトラックやシリンダ等のデータ単位に対して繰り返し更新を行なう場合もある。この様なボリュームデータの更新形態においては、同期式の二重化方式とした場合，中央処理装置からのデータ更新回数と同一の回数だけ第２のディスクサブシステムへのＷＲＴ＿Ｉ／Ｏが必要となる。一方、非同期方式の二重化方式とした場合、ある期間第１のディスク制御装置に未反映データを滞留させることによって、二重化を図る際の最新データのみをまとめて反映させれば良いため、第２のディスクサブシステムへのＷＲＴ＿Ｉ／Ｏの発行回数を、第１の中央処理装置からの第１のディスクサブシステムへのＷＲＴ＿Ｉ／Ｏ回数より削減することが可能である。非同期方式の二重化方式においては、いかに効率よく二重化データをまとめるかが性能向上の最大のポイントとなる。必要以上に第１のディスクサブシステム内に未反映データを滞留させることは、逆にキャッシュメ

モリの利用効率を下げ、性能劣化の要因となり得るからである。

【００１６】また、ＲＡＩＤ－５の記憶方式の場合、前述の全ストライプ方式の冗長データ作成方式とリードモディファイライト方式の冗長データ作成方式とでは、明らかに冗長データの作成効率に差が生じる。即ち、全ストライプ方式の冗長データ作成方式の場合、ｎ個の更新データに対し、ｍ回の冗長データの更新を行なうのに対し、リードモディファイライト方式の場合には１回の更新に対しｍ回の冗長データの更新が発生するからである。このため、できるだけ全ストライプライト方式の冗長データ作成を行なう方が、サブシステム全体のスループットを向上させることに繋がる。第１および第２のディスクサブシステムがＲＡＩＤ－５の記憶方式である場合、第２のディスクサブシステムへのＷＲＴ＿Ｉ／Ｏを起動する第１のディスク制御装置においては、自サブシステムの冗長データの作成効率を向上させるばかりでなく、第２のディスク制御装置が効率よく冗長データを作成可能なように二重化のＷＲＴ＿Ｉ／Ｏを発行することが、第２のディスク制御装置のスループットを向上させ、遠隔二重化システム全体のスループット向上に繋がる。

【００１７】本発明の目的は、稼働状況に応じてデータ複写の実行方法および契機を制御することにより、複数の記憶サブシステム間でのデータ多重化のためのデータ複写に伴う負荷の増大を抑制して各記憶サブシステムの処理性能を向上させることが可能なデータ多重化記憶サブシステムを提供することにある。

【００１８】本発明の他の目的は、稼働状況に応じてデータ複写の実行方法および契機を制御することにより、複数の記憶サブシステム間に設けられたデータ多重化のためのデータ転送経路の負荷の増大を抑制してデータ転送経路の使用効率を向上させることが可能なデータ多重化記憶サブシステムを提供することにある。

【００１９】本発明の他の目的は、多重化未完のデータ量の増大を抑止しつつ、複数の記憶サブシステム間でのデータ多重化のためのデータ複写の実行契機の最適化による性能向上を実現することが可能なデータ多重化記憶サブシステムを提供することにある。

【００２０】本発明の他の目的は、ＲＡＩＤ等の冗長記憶構成の複数の記憶装サブシステム間でのデータ多重化のためのデータ複写に伴う負荷の増大を抑制して各記憶サブシステムの処理性能を向上させることが可能なデータ多重化記憶サブシステムを提供することにある。

【００２１】

【課題を解決するための手段】本発明は、第１のデータ転送経路を介して上位装置と接続される第１の記憶サブシステムと、少なくとも一つの第２の記憶サブシステムとを第２のデータ転送経路にて接続し、第１の記憶サブシステムが上位装置から受領した書き込みデータを第２のデータ転送経路を介して第２の記憶サブシステムに複写することによりデータ多重化を行うシステムにおいて、たとえば第１の記憶サブシステム内の記憶制御装置に、配下の記憶装置におけるデータの所望の管理単位（たとえばトラック）、もしくは複数個の管理単位（たとえばシリンダ等）、またはボリューム単位、ファイル単位毎に一定期間内のデータ更新回数等を記憶するデータ更新回数記憶テーブルを制御情報記憶手段として持つ。

【００２２】このテーブルは、たとえば、ｎ世代前の記録までを保持できるようにｎ面のテーブル面を持つ。上位装置からのデータ更新時には、第１の記憶サブシステムの記憶制御装置の制御プログラム（制御論理）は、最新のデータ更新回数記憶テーブルの当該領域をカウントアップする。このカウントされた値は、当該領域内の第２の記憶サブシステムへの反映すべきデータの溜り具合を示す指標となる。また、一定周期毎にデータ更新回数記憶テーブルのデータをバックアップ化するとともに最新のデータ更新回数記憶テーブルをクリアする。また別の一定周期毎に、第２の記憶サブシステムへの未反映データを検索し、多重化のためのＷＲＴ＿Ｉ／Ｏ発行契機に、このデータ更新回数記憶テーブルを参照し、ｎ世代前からの更新回数の変化を調べ、更新回数の減少傾向にあるデータ領域を上位装置からのアクセスが終了しつつあるデータ領域と判断して、増加傾向にあるデータより優先的にサブシステム間のデータ複写をスケジュールするように制御する。一定周期毎の世代管理を行なうことによって、第２の記憶サブシステムへの未反映データの溜り具合の変化を捉えることが可能となる。

【００２３】さらに、第１の記憶サブシステムの記憶制御装置の制御プログラムは、たとえば、複数の記憶サブシステム間の全データの初期コピー処理による多重化の実行時に、当該多重化の対象範囲（たとえばボリューム）のデータ更新回数記憶テーブルを参照する。第１の記憶サブシステムの配下のマスタボリュームからのデータの呼び出しおよび第２の記憶サブシステムへのＷＲＴ＿Ｉ／Ｏの発行に際しては、このデータ更新回数記憶テーブルの値と世代間の差による増加・減少傾向から、以降に多重化の対象となる領域の今後の上位装置からの更新を予測し、まだ引き続き更新が継続するようであれば、スケジュールを遅らせる、等の最適化を行う。この様な、データ更新回数の記憶手段とデータ更新の継続性の推測手段によって、初期コピー処理の効率向上を図る。

【００２４】一方、各記憶サブシステムの記憶装置が単位データ群と、このデータ群から生成される冗長データを異なる記憶媒体に分散して格納する冗長記憶構成を備えている場合、前記データ更新回数記憶テーブルによる更新履歴から、冗長データを作成するストライプ列に対する更新の継続性を推測する手段を設ける。当該ストラ

イプ列への上位装置からの更新が継続するようであれば、多重化のための第２の記憶サブシステムへのＷＲＴ＿Ｉ／Ｏの発行スケジュールを遅らせ、全ストライプライト可能な更新データがそろってから、または、更新されていないストライプ列内データをも併せ、ストライプ列全体のデータを纏めて転送する。また、当該ストライプ列への更新の継続が見られない場合には、更新部分のみを転送する。この様な手段によって、第２の記憶サブシステムにおいて、第１の記憶サブシステムから到来する複写データの格納時における全ストライプライトとリードモディファイライトが効率良く制御可能なようにする。

【００２５】さらに、各記憶サブシステムの記憶装置が冗長記憶構成を備えている場合、多重化のためのＷＲＴ＿Ｉ／Ｏの完遂に掛かるデータ転送時間を含むＩ／Ｏ処理時間を観測・記憶する手段を設けることもできる。また、全ストライプ方式による冗長データの作成オーバヘッドとリードモディファイライト方式による冗長データの作成オーバヘッドを観測・記憶する手段を設ける。この観測されたＩ／Ｏ処理時間からストライプ列全体のデータ転送に掛かる時間を推測し、リードモディファイライト方式と全ストライプ方式との処理の時間差を比較し、よりオーバーヘッドの少ない方を選択する。

【００２６】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【００２７】（実施の形態１）図１は、本発明の一実施の形態であるデータ多重化記憶サブシステムの構成の一例を示す概念図、図２は、本実施の形態のデータ多重化記憶サブシステムにて用いられる制御情報の一例を示す概念図、図３および図４は、本実施の形態のデータ多重化記憶サブシステムの作用の一例を示すフローチャートである。

【００２８】まず、図１を用いて本実施の形態のデータ多重化記憶サブシステムの構成例を説明する。中央処理装置１０１から１ないしはｎ本のデータ転送パス１０２で接続されたディスクサブシステムをマスタディスクサブシステム１０４とし、二重化データを格納するバックアップ側のディスクサブシステムをリモートディスクサブシステム１０５とする。リモートディスクサブシステム１０５に接続するバックアップ用ＣＰＵをセカンダリＣＰＵとする。それぞれのディスクサブシステムは、ディスク制御装置１０３と、配下の１ないしｎ個のディスク記憶装置１１６とから構成される。このｎ個のディスク記憶装置１１６への記憶方式は前述のＲＡＩＤ－５であっても良い。ディスク制御装置１０３は、１ないしはｎ個の対ＣＰＵ制御用プロセッサ１０７と、１ないしはｎ個の対リモート制御用プロセッサ１１３、１ないしはｎ個の対ディスク制御用プロセッサ１１４を持つマルチプロセッサにより構成され、それぞれ、対ＣＰＵデータ転送ポート１０６、対リモートデータ転送ポート１１２、対ディスクデータ転送ポート１１５を介して外部とのデータ転送を制御する。この対ＣＰＵ制御用プロセッサ１０７と対リモート制御用プロセッサ１１３は同一のプロセッサでタイムシェアリング的に制御を切り替えるものであっても良い。また、各プロセッサから共通アクセス可能なキャッシュメモリ１１１と、各プロセッサの共通制御情報を格納する共通メモリ１０８を持ち、各プロセッサからは共通バス１１０にてアクセスされる。

【００２９】本実施の形態の場合には、この共通メモリ１０８に、データ更新回数記憶テーブル１０９を持つ。この様なマスタディスクサブシステム１０４の構成と同一の構成を持つディスクサブシステムを、たとえば遠隔地に設置してリモートディスクサブシステム１０５とし、各々のディスク制御装置１０３の間を１ないしはｎ本のマスターリモート間データ転送パス１１７にて接続する。このマスターリモート間データ転送パス１１７としては、たとえば専用通信回線や公衆通信回線等の任意の情報通信媒体や情報ネットワークを用いることができる。

【００３０】次に、中央処理装置１０１（プライマリＣＰＵ）からのマスタディスクサブシステム１０４へのデータ更新が発生したときの動作例について示す。図２は、データ更新回数記憶テーブル１０９の構成例である。データ更新回数記憶テーブル１０９は、二重化対象データの管理単位毎にエントリを持ち（本実施例ではシリンダ（ＣＹＬ）とする）、二重化対象の全領域分のデータを保持する。更に、複数世代にわたってデータ更新回数を管理する場合には、このデータ更新回数記憶テーブル１０９をｎ世代分（ｎ面；本実施例では３面、Ａ面～Ｃ面（２０１～２０３））作成・保持する。最新世代からｎ世代前までのデータを管理するために、世代管理ポインタテーブル２０４を持つ。世代管理ポインタテーブル２０４は、最新世代更新回数記憶テーブルポインタ２０４ａ、一世代前更新回数記憶テーブルポインタ２０４ｂ、二世代前更新回数記憶テーブルポインタ２０４ｃからなる。

【００３１】次に、上位の中央処理装置１０１すなわちプライマリＣＰＵからマスタディスクサブシステム１０４に対するＷＲＴ＿Ｉ／Ｏを受領したときの対ＣＰＵ制御用プロセッサ１０７の制御プログラムの制御（第１の制御動作）について、図３のフローチャートを用いて説明する。ステップ３０１で上位の中央処理装置１０１からのＷＲＴコマンドを受領すると、ステップ３０２でキャッシュメモリ１１１上のＷＲＴデータ格納領域を確保し、ステップ３０３で対ＣＰＵデータ転送ポート１０６にＣＰＵ－キャッシュメモリ間のデータ転送を起動し、ステップ３０４でハードウェアのデータ転送完了を待つ。データ転送が完了するとステップ３０５で世代管理ポインタテーブル２０４から最新世代更新回数記憶テー

ブルポインタ２０４ａに対応した最新世代の更新回数記憶テーブルＡ面（２０１）を得て、ステップ３０６で当該更新部分に対応する最新のデータ更新回数記憶テーブル１０９のエントリをカウントアップし、ステップ３０７で上位の中央処理装置１０１へＷＲＴ＿Ｉ／Ｏの終了報告を行なう。当該マスタディスクサブシステム１０４配下のディスク記憶装置１１６への実際の書き込みは、本制御を行なっている対ＣＰＵ制御用プロセッサ１０７とは異なる対ディスク制御用プロセッサ１１４の制御によって非同期に書き込まれるものであっても良い。この様に、上位の中央処理装置１０１からのＷＲＴコマンド受領時に、最新世代のデータ更新回数記憶テーブル１０９の対応する領域をカウントアップすることによって、データ更新回数の履歴を記憶する。

【００３２】次に、対リモート制御用プロセッサ１１３のデータ二重化のためのＷＲＴ＿Ｉ／Ｏ発行処理に関する制御プログラムの制御の一例について、図４および図５を用いて説明する。図４は主にデータ更新回数記憶テーブル１０９の制御に関する処理フローの例である。対リモート制御用プロセッサ１１３の制御プログラムは、ダイナミックにループしながら特定周期毎にデータ更新回数記憶テーブル１０９の管理とリモートディスクサブシステム１０５へのＷＲＴ＿Ｉ／ＯのスケジュールおよびＷＲＴ＿Ｉ／Ｏ実行処理を行なう。ここでステップ４０１の周期Ａは、データ更新回数記憶テーブル１０９の制御を周期的に行なう処理の起動周期であり、ステップ４０２の周期Ｂは二重化ＷＲＴ＿Ｉ／Ｏスケジュールおよび実行処理の起動周期である。周期Ａと周期Ｂは、Ａ＞Ｂの関係の適当な周期とする。ステップ４０１で周期Ａの経過を検知すると、データ更新回数記憶テーブル１０９の管理の処理を行なう。すなわち、ステップ４０３で、世代管理ポインタテーブル２０４の二世代前のデータ更新回数記憶テーブル１０９を指すポインタテーブル（二世代前更新回数記憶テーブルポインタ２０４ｃ）の値を一時的にワークエリアに退避し、ステップ４０４で一世代前のデータ更新回数記憶テーブル１０９を指すポインタ値（一世代前更新回数記憶テーブルポインタ２０４ｂ）を二世代前のデータ更新回数記憶テーブル１０９を指すポインタ格納領域（二世代前更新回数記憶テーブルポインタ２０４ｃ）に複写する。ステップ４０５で最新のデータ更新回数記憶テーブル１０９を指すポインタ値（最新世代更新回数記憶テーブルポインタ２０４ａ）を一世代前の更新回数記憶テーブル１０９を指すポインタ（一世代前更新回数記憶テーブルポインタ２０４ｂ）に複写する。ステップ４０６でワークエリアに退避してあった元の二世代前の更新回数記憶テーブルへのポインタ値を最新世代更新回数記憶テーブルポインタ２０４ａの格納領域に複写し、ステップ４０７で最新世代のデータ更新回数記憶テーブル１０９となったテーブル面をクリアし最新化を図る。この様にして、特定周期でデータ更新回数記憶テーブル１０９の複数面を入れ替え世代管理を実現する。ステップ４０２で周期Ｂの経過を検知すると二重化ＷＲＴ＿Ｉ／Ｏスケジュールおよび実行処理を行なう（ステップ４１０）。

【００３３】この二重化ＷＲＴ＿Ｉ／Ｏスケジュールおよび実行処理の一例を示すフローチャートを図５に示す。ステップ５０１からステップ５０５がマスタディスクサブシステム１０４上に溜まっているリモートディスクサブシステム１０５への未反映データの検索処理である。この処理において、本実施の形態にて例示されるデータ更新回数記憶テーブル１０９等の手段によって、更新回数の時系的な変化を捉え、ＷＲＴ＿Ｉ／Ｏ実行の対象とするか否かの判断を行なう。具体的には、ステップ５０１でキャッシュメモリ１１１上のリモートディスクサブシステム１０５への未反映データを検索する。この検索は、たとえばハッシュを用いたキャッシュメモリ１１１上のデータ管理方法によるものであっても良いし、ＬＲＵアルゴリズムによるものであっても良い。ステップ５０２で検索された未反映データの領域に対応する更新回数を、各世代毎のデータ更新回数記憶テーブル１０９から読み出す。ステップ５０３で、二世代前の更新回数から一世代、最新世代と比較し、増減傾向を調べる。比較結果から当該領域への更新が増加傾向にあると判断できる場合には、この時点での当該データのスケジュールを見送り別の未反映データの検索処理を行なう。また、全未反映データの検索が終了した場合には、もとの処理に戻る（ステップ５０４～５０５）。ステップ５０４で当該未反映データの範囲に対する更新が増加傾向に無い場合は、当該領域を二重化ＷＲＴの対象とし、リモートディスクサブシステム１０５へＷＲＴ＿Ｉ／Ｏを発行する（ステップ５０６～５１０）。

【００３４】次に、この二重化ＷＲＴ＿Ｉ／Ｏ実行処理について一例を記す（ステップ５０６～５１０）。

【００３５】まず、当該未反映データに対応するリモートディスクサブシステム１０５のデバイスに対しＷＲＴコマンドを発行する（ステップ５０６）。

【００３６】ステップ５０７で当該ＷＲＴコマンドに伴うデータ転送の完了待ちを行い、転送完了後、ステップ５０８でリモートディスクサブシステム１０５からのＩ／Ｏ完了報告待ちを行う。そしてステップ５０９にてエラー判定を行い、このＩ／Ｏ完了報告が異常終了であった場合には、ステップ５１０のエラー処理を行い、正常終了であった場合には元の周期監視処理に戻る。

【００３７】このように、ランダム性に富んだ上位の中央処理装置１０１からのディスクデータの更新に際しては、そのままでは正確な予測は不可能であるが、本実施の形態のデータ多重化記憶サブシステムによれば、たとえば、多重化のためのＷＲＴ＿Ｉ／Ｏ発行契機に、データ更新回数記憶テーブル１０９を参照し、ｎ世代前からの更新回数の変化を調べ、更新回数の減少傾向にあるデ

ータ領域を上位の中央処理装置１０１からのアクセスが終了しつつあるデータ領域と判断して、増加傾向にあるデータより優先的にマスターリモート間データ転送パス１１７を経由したサブシステム間のデータ複写をスケジュールするように制御することで、各データ領域における更新状況に応じた必要最小限のデータ複写回数にてデータの二重化を効率よく達成することが可能となる。また、一定周期毎の世代管理を行なうことによって、リモートディスクサブシステム１０５への未反映データのキャッシュメモリ１１１等における溜り具合の変化を捉えることが可能となり、データ複写の遅延に起因するキャッシュメモリ１１１等の利用効率に低下を回避して、システムのキャッシュメモリ１１１等の資源の可用性の向上を実現できる。

【００３８】さらに、データ複写のためのマスターリモート間データ転送パス１１７を経由した無駄なデータ転送が減り、マスターリモート間データ転送パス１１７を構成する情報通信媒体や情報ネットワーク等の負荷（トラヒック）の増大を防止して、情報通信媒体や情報ネットワーク等の効率的な利用によるデータ多重化が可能になる。

【００３９】（実施の形態２）次に、図１に例示される構成の本発明のデータ多重化記憶サブシステムにおいて、マスタディスクサブシステム１０４とリモートディスクサブシステム１０５との間における初期コピー操作等における効率的なデータ複写の実現方法（第２の制御動作）の一例を、図６のフローチャートにて説明する。

【００４０】たとえば、図１に示すマスタディスクサブシステム１０４上のボリュームデータを新たに二重化ペアとして設定する場合、マスタディスクサブシステム１０４のディスク制御装置１０３は、自配下の当該ボリュームデータをキャッシュメモリ１１１上にステージングし、そのデータをＷＲＴ＿Ｉ／Ｏによってリモートディスクサブシステム１０５へ書き込む。この動作を当該ボリュームの全トラック（ＴＲＫ）範囲に渡って繰り返すことによって初期コピー（初期のボリュームデータ多重化）を行なう。自配下のボリュームからのステージング処理は図１に示す対ディスク制御用プロセッサ１１４が制御し、リモートディスクサブシステム１０５へのＷＲＴ＿Ｉ／Ｏの発行は対リモート制御用プロセッサ１１３が制御する。

【００４１】図６に例示されるフローチャートは、初期のボリュームデータ多重化にて複写対象領域に対する上位の中央処理装置１０１からのデータ更新要求の有無や頻度等に応じてデータ複写の順序を動的に変更する機能を持った対リモート制御用プロセッサ１１３の初期コピー処理例である。ここで、本実施の形態に示すデータ更新回数記憶テーブル１０９の世代管理については、先の図２および図３に例示した場合と同様の制御を行う。対リモート制御用プロセッサ１１３は、初期コピー処理においては、初期コピー対象のデバイスを選択し（ステップ６０１）、対象のデバイスの次コピー対象領域の各世代毎のデータ更新回数記憶テーブル１０９の値を読み出す（ステップ６０２）。ここで１回のリモートディスクサブシステム１０５へのコピーの単位は複数のデータ単位（トラック）であるシリンダ単位であっても良いし、またトラック単位であっても良い。ステップ６０３で各世代毎の更新回数を比較し、もし次コピー対象領域に対する上位からの更新が増加傾向にある場合には、当該ボリュームの今回のコピー処理を見合わせ、もし初期コピー処理が複数個のボリュームで同時になされている場合には、別の初期コピー対象のデバイスの初期コピー処理のスケジュールを行なう（ステップ６０４およびステップ６１３）。ステップ６０４で次コピー対象領域への更新が増加傾向にない場合は、当該領域を次コピー対象領域と定めコピー対象範囲内のトラックがキャッシュメモリ１１１上に存在する（Ｃａｃｈｅ　Ｈｉｔ）か存在しないか（Ｃａｃｈｅ　Ｍｉｓｓ）を検索する（ステップ６０５）。ステップ６０６でＣａｃｈｅ　Ｍｉｓｓであるトラックのステージング処理要求を対ディスク制御用プロセッサ１１４に発行し、ステップ６０７でそのステージング処理完了待ちを行なう。コピー対象領域のステージング処理が完了し、全トラックがキャッシュメモリ１１１上にステージングされている状態で二重化のためのＷＲＴ＿Ｉ／Ｏ発行をリモートディスクサブシステム１０５に行なう（ステップ６０８から６１２）。この処理は、先に図５に例示した実施例の処理と同一である。ステップ６１１にて二重化のためのＷＲＴ＿Ｉ／Ｏ処理が正常に終了すると、ステップ６１３で別の初期コピー対象デバイスを選択し、これまでと同様の制御を繰り返し、初期コピー処理を完成させる。

【００４２】このように、本実施の形態の場合には、たとえばシステム立ち上げ時等の契機にて実行される、マスタディスクサブシステム１０４の配下のディスク記憶装置１１６の全データをリモートディスクサブシステム１０５に複写する初期データ複写等において、データ複写中に複写予定のデータ領域に上位からの更新要求が集中するような場合には、当該更新要求が無くなるまで当該領域の複写を後回しにして、他の更新要求が発生していない安定なデータ領域の複写を先行させる等の制御を行うことで、初期データ複写等の処理を効率よく遂行することが可能になる。

【００４３】（実施の形態３）次に、図１に例示される構成のデータ多重化記憶サブシステムにおいて、ＲＡＩＤ方式にて、マスタディスクサブシステム１０４およびリモートディスクサブシステム１０５にてデータを格納する場合のデータ二重化のためのデータ複写の効率化の一例について説明する。

【００４４】図７に本実施の形態におけるＲＡＩＤ方式のデータ格納の例を示す。ＲＡＩＤ方式によるデータの

記憶は、論理ボリュームをｎ＋１個のデバイスで構成し、データの格納単位（たとえばトラック）毎に、デバイスを分けて格納する。データ（トラック）はｎ個の単位でストライプを成し、それに対し一つまたは複数の冗長データを持つ。図７に示す格納例では、トラック番号０からｎ－１までのトラックを１ストライプとし一つの冗長データ（Ｐａｒｉｔｙ＃０）を持つ。同様にトラック＃ｎ×１から＃ｎ×１＋（ｎ－１）までのストライプに対しＰａｒｉｔｙ＃１を持つ。ここで冗長データの配置は常に固定のデバイスに配置する方式であっても良いし、図７の例に示すように順次に格納デバイスを変えて格納する方式であっても良い。ＲＡＩＤ方式におけるデータの更新は、前述の様に全ストライプ列で冗長データを作成する全ストライプライト方式と、更新データとその旧データおよび旧冗長データから新冗長データを作成するリードモディファイライト方式とがある。冗長データを構成するストライプ列中の更新データが複数個に渡る場合には、全ストライプライト方式で冗長データを作成する方がそのオーバヘッドは削減される。

【００４５】まず、特定のストライプ内の各データ単位（この場合はトラック）の各々に対する更新要求の発生状況に応じて、複写先のリモートディスクサブシステム１０５に対して、更新されたトラックのデータのみを転送してリードモディファイライト方式でのデータ格納を促す（第２の複写操作）か、更新トラックを含む全ストライプデータの転送によって、全ストライプライト方式でのデータ格納を促す（第１の複写操作）かを動的に切り換える場合（第３の制御動作）について説明する。

【００４６】すなわち、二重化のためのＷＲＴ＿Ｉ／Ｏ発行対象のデータのストライプ範囲の更新回数の世代毎の更新履歴から増加傾向を判断し、増加傾向にある場合には当該ストライプ列の更新が継続するとの判断から、ストライプ列での更新を促すようにストライプ列範囲全般に渡って更新がなされるまで、ＷＲＴ＿Ｉ／Ｏの発行を遅らせる。減少傾向にある場合、または更新の継続性が見られない場合には、当該データをスケジュールした時点で二重化のためのＷＲＴ＿Ｉ／Ｏの発行を行なう。さらに、この場合更新された部分のみを二重化のためのＷＲＴ＿Ｉ／Ｏで転送してもよいし、更新はされていない同一ストライプ列の他のデータを併せてＷＲＴ＿Ｉ／Ｏでリモートディスクサブシステム１０５に書き込むことによって、リモートディスクサブシステム１０５にて全ストライプライトが促進されるように制御することも可能である。

【００４７】上述の例では、ストライプ内の各トラックに対する更新要求の有無にてストライプ内の全トラックをまとめて転送するか、更新のあったトラックのみを転送するかを切り換えていたが、さらに、この切り換えの判定条件として、マスタディスクサブシステム１０４からリモートディスクサブシステム１０５へのデータ転送時間を含めたデータ複写の全所要時間の大小に応じて、リードモディファイライト方式か、全ストライプライト方式かを選択させる場合（第４の制御動作）の一例について、図８および図９を参照して以下に説明する。データ転送の遅延時間は、５ｎｓ／ｍなので、複数の記憶サブシステムがたとえば数百キロも離れた遠隔地に設置された場合、両者間におけるデータ転送所要時間は各サブシステム内の処理時間に比較して無視できないほど大きくなり、このデータ転送所要時間を加味して、データ転送方法を切り換えることは、データ複写処理の効率改善において大きな意味を持つ。

【００４８】すなわち、この判断基準に二重化のためのＷＲＴ＿Ｉ／Ｏに要するトラック単位の処理オーバヘッドを観測する手段として、図１に示す共通メモリ１０８に、図８に例示される構成の平均ＷＲＴ＿Ｉ／Ｏ処理時間格納テーブル８０１と平均値観測回数カウンタ８０２（平均値観測回数：Ｎ）を設ける。

【００４９】図９にトラック単位の処理オーバヘッドの観測のフローチャートの一例を示す。平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）を観測する二重化のためのＷＲＴ＿Ｉ／Ｏ発行処理の基本的な流れは、図５に示すものと同様である。この図５に例示された処理の流れに加えて図９の例では、更に、ステップ９１１でデータ転送開始時刻を記憶し、ステップ９１４でＩ／Ｏ完了報告が正常になされたか否かを判定した後、ステップ９１６で現時刻とデータ転送開始時刻との差からＷＲＴ＿Ｉ／Ｏ処理時間（Δｔ）を算出する。ステップ９１７で、これまでの平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）と今回計測されたＷＲＴ＿Ｉ／Ｏ処理時間（Δｔ）とから最新の平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）を算出する。また、ステップ９１８で算出した最新の平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）を平均ＷＲＴ＿Ｉ／Ｏ処理時間格納テーブル８０１に格納し、ステップ９１９で平均値観測回数カウンタ８０２（Ｎ）をインクリメントする。尚、ステップ９１７の算出（平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）の更新）方法は、たとえば下記の（式１）に例示される通りである。

【００５０】

【数１】

$$t_m \Leftarrow \frac{t_m \times N + \Delta t}{N+1} \quad \text{(式1)}$$

【００５１】上記のようにして平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）を観測し、このＷＲＴ＿Ｉ／Ｏ処理オーバヘッドを元に全ストライプ列のデータを併せて、リモートディスクサブシステム１０５に更新を行なった方がトータルの処理時間が削減されるか否かを判断する。この判断ステップは、たとえば、図９におけるステップ９１０の直前に、平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）の推移に応じて、判定ルーチンを実行することで実現できる。たとえば、リードモディファイライト方式、および

全ストライプライト方式の各々において観測された平均ＷＲＴ＿Ｉ／Ｏ処理時間（ｔｍ）の大小を比較し、ｔｍがより小さい方式を選択してデータ複写を実行する、等の方法が考えられる。

【００５２】このように、マスターリモート間データ転送パス１１７におけるデータ転送時間を含めたデータ複写処理の全所要時間を観測することにより、マスタディスクサブシステム１０４およびリモートディスクサブシステム１０５にてＲＡＩＤ方式でデータを格納する場合において、データ複写先のリモートディスクサブシステム１０５でリードモディファイライト方式および全ストライプライト方式のいずれの方式でデータ格納動作を行わせるかをよりきめ細かく制御でき、データ多重化のためのデータ複写の最適化および効率化を実現することが可能になる。

【００５３】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【００５４】

【発明の効果】本発明のデータ多重化記憶サブシステムによれば、稼働状況に応じてデータ複写の実行方法および契機を制御することにより、複数の記憶サブシステム間でのデータ多重化のためのデータ複写に伴う負荷の増大を抑制して各記憶サブシステムの処理性能を向上させることができる、という効果が得られる。

【００５５】また、本発明の、データ多重化記憶サブシステムによれば稼働状況に応じてデータ複写の実行方法および契機を制御することにより、複数の記憶サブシステム間に設けられたデータ多重化のためのデータ転送経路の負荷の増大を抑制してデータ転送経路の使用効率を向上させることができる、という効果が得られる。

【００５６】また、本発明のデータ多重化記憶サブシステムによれば、多重化未完のデータ量の増大を抑止しつつ、複数の記憶サブシステム間でのデータ多重化のためのデータ複写の実行契機の最適化による性能向上を実現することができる、という効果が得られる。

【００５７】本発明のデータ多重化記憶サブシステムによれば、ＲＡＩＤ等の冗長記憶構成の複数の記憶装サブシステム間でのデータ多重化のためのデータ複写に伴う負荷の増大を抑制して各記憶サブシステムの処理性能を向上させることができる、という効果が得られる。

【図面の簡単な説明】

【図１】本発明の一実施の形態であるデータ多重化記憶サブシステムの構成の一例を示す概念図である。

【図２】本発明の一実施の形態であるデータ多重化記憶サブシステムにて用いられる制御情報の一例を示す概念図である。

【図３】本発明の一実施の形態であるデータ多重化記憶サブシステムの作用の一例を示すフローチャートである。

【図４】本発明の一実施の形態であるデータ多重化記憶サブシステムの作用の一例を示すフローチャートである。

【図５】本発明の一実施の形態であるデータ多重化記憶サブシステムの作用の一例を示すフローチャートである。

【図６】本発明の一実施の形態であるデータ多重化記憶サブシステムの作用の一例を示すフローチャートである。

【図７】本発明の一実施の形態であるデータ多重化記憶サブシステムにおけるＲＡＩＤ方式のデータ格納方法の一例を示す概念図である。

【図８】本発明の一実施の形態であるデータ多重化記憶サブシステムにて用いられる制御情報の一例を示す概念図である。

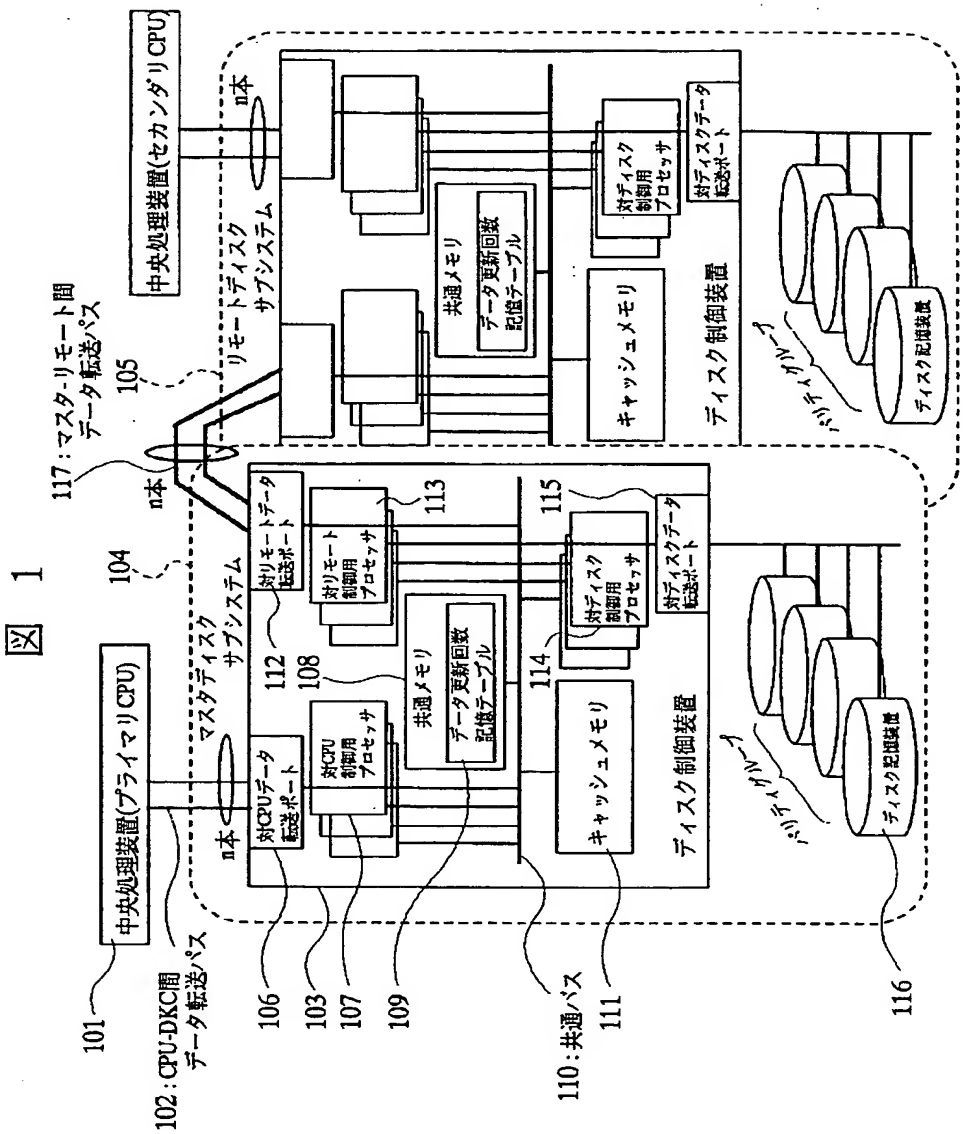
【図９】本発明の一実施の形態であるデータ多重化記憶サブシステムの作用の一例を示すフローチャートである。

【符号の説明】

１０１…中央処理装置（上位装置）、１０２…データ転送パス（第１のデータ転送経路）、１０３…ディスク制御装置、１０４…マスタディスクサブシステム（第１の記憶サブシステム）、１０５…リモートディスクサブシステム（第２の記憶サブシステム）、１０６…対ＣＰＵデータ転送ポート、１０７…対ＣＰＵ制御用プロセッサ、１０８…共通メモリ、１０９…データ更新回数記憶テーブル（制御情報記憶手段）、１１０…共通バス、１１１…キャッシュメモリ、１１２…対リモートデータ転送ポート、１１３…対リモート制御用プロセッサ、１１４…対ディスク制御用プロセッサ、１１５…対ディスクデータ転送ポート、１１６…ディスク記憶装置、１１７…マスターリモート間データ転送パス（第２のデータ転送経路）、２０４…世代管理ポインタテーブル、２０４ａ…最新世代更新回数記憶テーブルポインタ、２０４ｂ…一世代前更新回数記憶テーブルポインタ、２０４ｃ…二世代前更新回数記憶テーブルポインタ、８０１…平均ＷＲＴ＿Ｉ／Ｏ処理時間格納テーブル、８０２…平均値観測回数カウンタ。

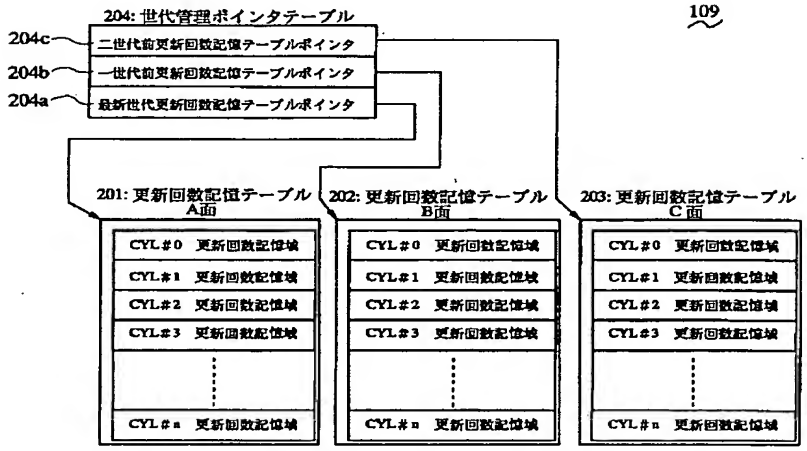
【図１】

図 1


中央処理装置(セカンダリCPU)
中央処理装置(プライマリCPU)
117:マスタ-リモート間 データ転送パス
105
104
n本
リモートディスク サブシステム
マスタディスク サブシステム
対リモートデータ転送ポート
対リモート制御用プロセッサ
対CPUデータ転送ポート
対CPU制御用プロセッサ
共通メモリ
データ更新回数記憶テーブル
対ディスク制御用プロセッサ
対ディスクデータ転送ポート
キャッシュメモリ
ディスク制御装置
パリティグループ
ディスク記憶装置
112
113
108
114
115
101
102:CPU-DKC間 データ転送パス
106
103
107
109
110:共通バス
111
116

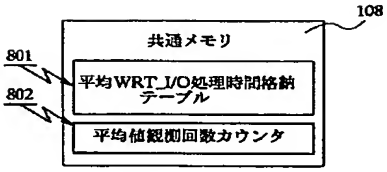
【図２】

図　２



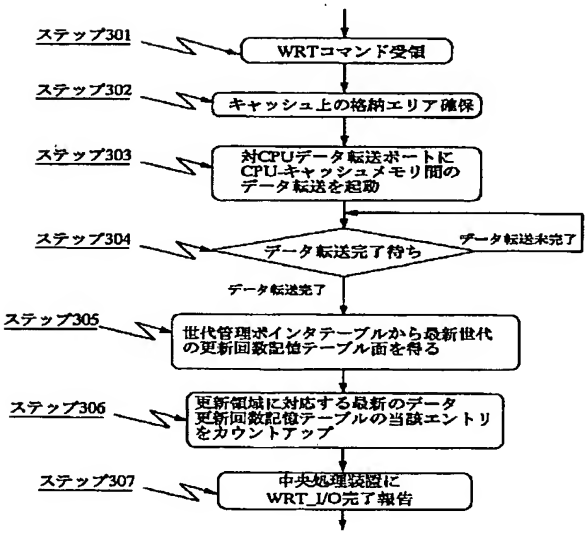

【図８】

図　８



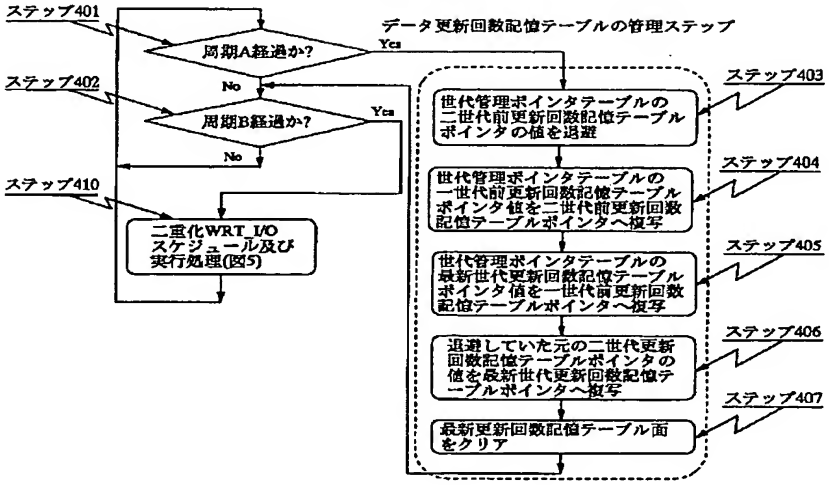

【図３】

図　３



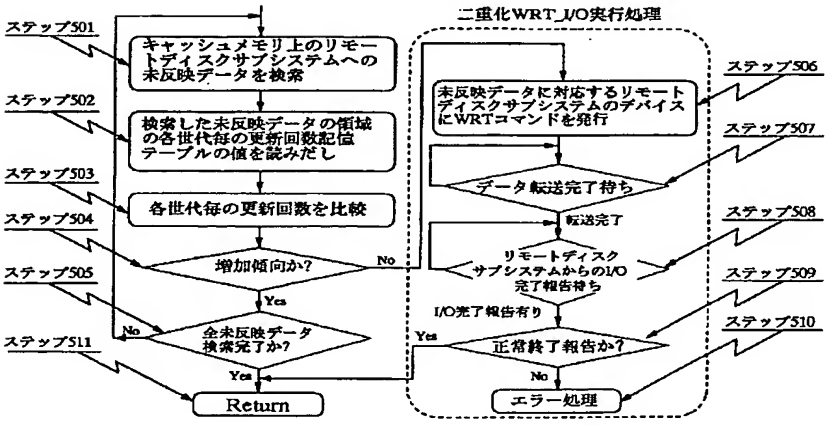
【図４】

図　4

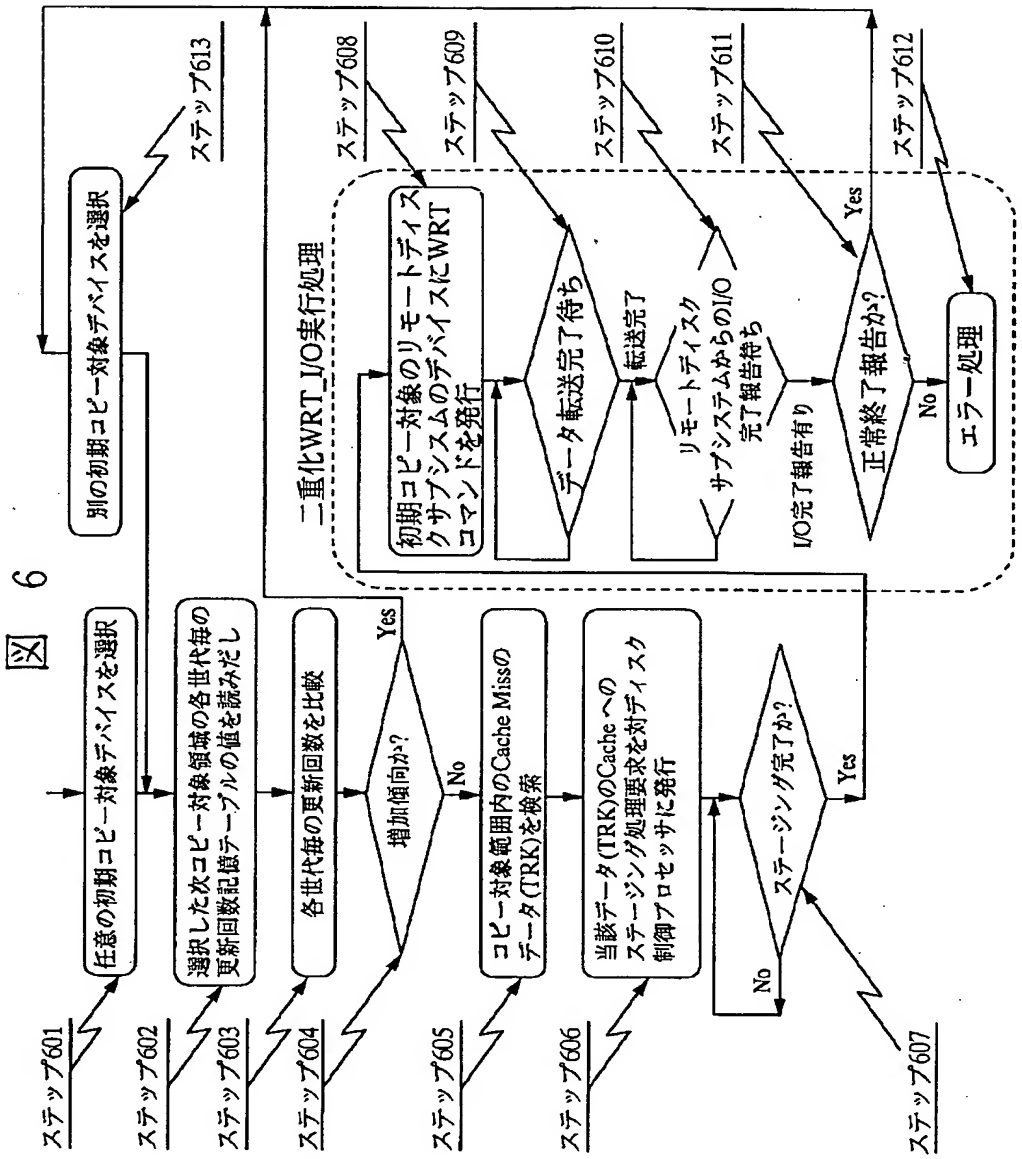

ステップ401
周期A経過か?
Yes
No
ステップ402
周期B経過か?
Yes
No
ステップ410
二重化WRT_I/O
スケジュール及び
実行処理(図5)
データ更新回数記憶テーブルの管理ステップ
ステップ403
世代管理ポインタテーブルの
二世代前更新回数記憶テーブル
ポインタの値を退避
ステップ404
世代管理ポインタテーブルの
一世代前更新回数記憶テーブル
ポインタ値を二世代前更新回数
記憶テーブルポインタへ複写
ステップ405
世代管理ポインタテーブルの
最新世代更新回数記憶テーブル
ポインタ値を一世代前更新回数
記憶テーブルポインタへ複写
ステップ406
退避していた元の二世代更新
回数記憶テーブルポインタの
値を最新世代更新回数記憶テ
ーブルポインタへ複写
ステップ407
最新更新回数記憶テーブル面
をクリア

【図５】

図　5

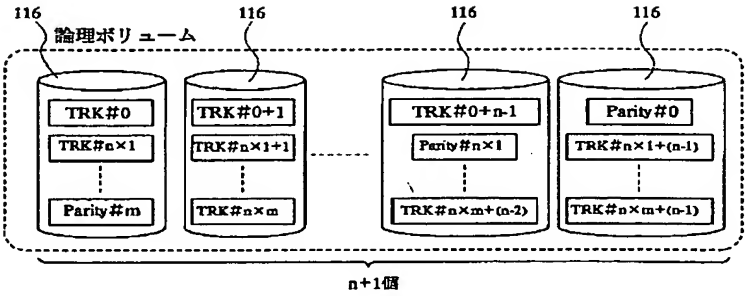

ステップ501
キャッシュメモリ上のリモート
ディスクサブシステムへの
未反映データを検索
ステップ502
検索した未反映データの領域
の各世代毎の更新回数記憶
テーブルの値を読みだし
ステップ503
各世代毎の更新回数を比較
ステップ504
増加傾向か?
No
Yes
ステップ505
全未反映データ
検索完了か?
No
Yes
ステップ511
Return
二重化WRT_I/O実行処理
ステップ506
未反映データに対応するリモート
ディスクサブシステムのデバイス
にWRTコマンドを発行
ステップ507
データ転送完了待ち
転送完了
ステップ508
リモートディスク
サブシステムからのI/O
完了報告待ち
I/O完了報告有り
ステップ509
Yes
正常終了報告か?
No
ステップ510
エラー処理

【図６】



【図７】

図　7




【図９】

図　9